**Assignment: Assessing Airline Performance through Sentiment Analysis**

Choose an airline to analyze the sentiment of for your assignment from one of the following:

* Alaska
* Delta
* United

Follow the instructions in the file *airline-text-analysis.Rmd* to answer the following questions. Note that some of the questions are inside of the Rmd file, but you can answer them here using the output from your code in the Rmd file. To access this file, follow the instructions in *Getting Started in RStudio Cloud.docx*.

**Part 1**

Use the function make\_wordcloud() to make a wordcloud with the airline of your choice. Compare your wordcloud to American Airlines (below):

Text

Description automatically generated

1. Are the most frequent words in two wordclouds different?
   1. If they are not different, what can you infer from that?
   2. If they are different, what can you infer from that?
2. Can you piece together a story about each wordcloud based on the frequently used words?
3. Did anything surprise you?
4. What will you tell the marketing managers of these airlines that will give them new insights about their fliers?

**Part 2**

Consider the summary sentiment for JetBlue and note that we see that on average there are 3.76% positive words in a tweet compared to only 1.883% negative words.

To find this, check out the printed output from the code chunk:

Chart

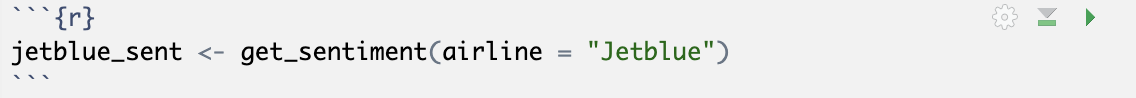
Description automatically generated

Execute the next code chunk to get a similar output for your airline of choice.

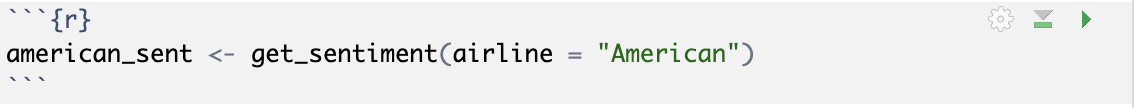
1. What percent of positive words are there in a tweet for your airline of choice, on average?
2. What percent of negative words are there in a tweet for your airline of choice, on average?
3. How does this compare to Jetblue?

**Part 3**

Create a sentiment data set for another airline of your choice. Note that to do this, you’ll need to alter the code below to use your airline of choice instead of Jetblue.



You may also want to rename your variable so it’s less confusing. Here is an example using American Airlines:



Now you’ll generate sentiment trends for your chosen airline and JetBlue. Here is a lineplot Jetblue’s negative tweets:

Chart, line chart

Description automatically generated

You can create one for your airline of choice by using the make\_lineplot() function. Compare the lineplot for your airline of choice and Jetblue (below), using *negative* emotion.

Here is an example using the American Airlines data we generated above:

Text

Description automatically generated with medium confidence

1. How does the airline you chose compare to Jetblue in terms of negative tweets?
2. Change the emotion for each to positive. How does the airline you chose compare to Jetblue’s positive tweets?
3. Based on these graphs, do you think customers had a better time flying with Jetblue in late July or in early June?
4. Based on these graphs, do you think customers had a more positive experience with Jetblue or your airline of choice? Why?

**Part 4**

Finally, use the make\_barplot() function to generate frequencies of each sentiment analyzed for each airline.

1. If you wanted to have the most *positive* experience, which 2 airlines might you consider based on these graphs?
2. Which airline seems to have faced the most *angry* tweets this summer?
3. Consider the following for your airline of choice:
   1. Ultimately, do their customers seem positive, trusting, and full of joy relative to other airlines?
   2. How frequent are negative sentiments (e.g. negative, fear, sadness) tweeted to your airline of choice relative to others?

How do customer sentiments for your airline of choice compare to the others? Use the above questions as a framework to give feedback to this airline. Write a paragraph explaining how are they doing and any areas you think they could improve based on this sentiment analysis.

**Part 6: Cleaning Text**

Consider a single tweet, such as the one you can print using:

as.character(airlinedata[[1]][2])

1. How might this data be cleaned for sentiment analysis?
2. Look at the file *data-sets-creation.R* and determine which stop words are being used. How many words are there?
3. Choose 3 stopwords that aree not part of the stopwords list and insert code to remove those as well. Add this to the file *data-sets-creation.R*.

**Part 7: Making Visualizations**

Look for the function make\_lineplot().

1. Instead of using the function make\_lineplot(), generate your own line graph for the sentiment *anger* with the variable jetblue\_sent using the package *ggplot2*.
2. How else could you visualize this data?